Washington Law Review

Volume 89
Number 4 *Symposium: Compensated Surrogacy in the Age of Windsor*

12-1-2014

Promoting Innovation While Preventing Discrimination: Policy Goals for the Scored Society

Frank Pasquale

Danielle Keats Citron

PROMOTING INNOVATION WHILE PREVENTING DISCRIMINATION: POLICY GOALS FOR THE SCORED SOCIETY

Frank Pasquale[*] & Danielle Keats Citron[**]

Professor Zarsky's response[1] is an erudite and thoughtful analysis of the discrimination concerns raised by our article, *The Scored Society*.[2] We particularly appreciate his connection of themes in our article with literature on discrimination law. This historical awareness and theoretical sophistication demonstrates the deep continuity between our concerns and those of other legal scholars.

Professor Zarsky has led us to realize that there are in fact several normative theories of jurisprudence supporting our critique of the scored society, which complement the social theory and political economy presented in our article. In this response, we clarify our antidiscrimination argument while showing that is only one of many bases for the critique of scoring practices. The concerns raised by Big Data may exceed the capacity of extant legal doctrines. Addressing the potential injustice may require the hard work of legal reform.

Before responding, though, we should acknowledge Professor Zarsky's contributions to the field, and explain how we believe our work advances inquiry along some of the trails he has blazed with his insightful analyses of data mining, privacy, and information law generally.

Professor Zarsky has done a great deal to explore the legal problems

---

[*] Professor of Law, University of Maryland Francis King Carey School of Law, Affiliate Fellow, Yale Information Society Project.

[**] Lois K. Macht Research Professor & Professor of Law, University of Maryland Francis King Carey School of Law, Affiliate Fellow, Yale Information Society Project, Affiliate Scholar, Stanford Center on Internet & Society. The authors are grateful to Professor Tal Zarsky for his thoughtful reply and innovative work, to Mallory Gitt, Maureen Johnston, Jessica Knowles, and their colleagues at the *Washington Law Review* for their terrific editing, and to Professor Ryan Calo for his insights and support.

1. Tal Z. Zarsky, *Understanding Discrimination in the Scored Society*, 89 WASH. L. REV. 1375 (2014).

2. Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014).

raised by the proliferation of networked identities and selves. We are all aware of the freedoms and dangers posed by pseudonyms, anonymous handles, and multiple roles online.[3] Viewed in light of Professor Zarsky's proposals about traceable anonymity to ensure accountability in a digital age, one can think of the scores explored in our article as deeply connected to those concerns: *quantified identities imposed on individuals*, often without their consultation, consent, or even awareness.

An attention economy has gradually developed on the internet, as companies collect information about the habits and demographics of those who visit their websites. Unlike the old broadcast model of simply exchanging programming content for (an easily avoided) obligation to watch commercials, the new online data collectors enjoy far greater powers to monitor the behavior and actions of users and to influence their online experience and reputation. The skillful use of that data is a large part of the success of online behemoths and is increasingly driving decisionmaking at companies ranging from banks to retailers.

But data collection and analysis raises serious concerns. Data collection practices range from the careful to the careless. The tradeoff between checking accuracy and speedy production can easily tilt toward the latter as competition increases.[4] Once primarily directed at marketing, data collection practices now figure into employment and credit opportunities.[5] The companies' digital stockpiles would delight a new Stasi or J. Edgar Hoover. Assurances to customers that data are anonymized mean little without audits—which are nonexistent for most firms, and rare and often cursory even when required. Should a company that observes a customer looking at a $10,000 ring on one site use that information to allow others to systematically raise prices for the customer on the assumption that he or she is wealthy?

What self-help measures can (and should) consumers take as they are observed online? Can contract and tort law address all of the potential violations of privacy that occur due to interferences with our settled expectations about how our data is used? Can law address the harms

---

3. DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 27–28, 57–62, 222–25 (2014) (exploring the promise and perils of online anonymity, including cyber harassment, and calling for anonymity as a default privilege that can be lost).

4. Natasha Singer, *When Your Data Wanders to Places You've Never Been*, N.Y. TIMES, Apr. 28, 2013, at BU3; Shannon Pettypiece & Jordan Robertson, *Did You Know You Had Diabetes? It's All Over the Internet*, BLOOMBERG (Sept. 11, 2014, 1:07 PM), http://www.bloomberg.com/news/2014-09-11/how-big-data-peers-inside-your-medicine-chest.html.

5. Frank Pasquale, Op-Ed., *The Dark Market for Personal Data*, N.Y. TIMES, Oct. 17, 2014, at A27.

associated with data leaks due to insecure systems?[6] The collection of vast reservoirs of data raises difficult questions across the range of public and private, and statutory and common, law.

Professor Zarsky's essay *Privacy and Data Collection in Virtual Worlds* was an early effort to tackle these problems.[7] The term "virtual world" is usually associated with game-like activity in venues like *Second Life* or *World of Warcraft*. But the metaphor of the virtual world helps enrich our understanding of the degrees of freedom available to the legal system as it addresses questions like privacy and identity online. Professor Zarsky highlighted the vast extent of personal information that can be collected in virtual worlds. He discussed the heightened level of surveillance prevalent in the online environment. In our era of pervasive surveillance, tracking, and the internet of things, that world is *our world*, and its most gifted explicators have given us important insights into how pervasive data-gathering and processing should be governed.[8]

While many laissez-faire commenters have claimed that users can "take or leave" participation in virtual worlds if they find such surveillance oppressive, Professor Zarsky early on realized how many important activities are migrating to these virtual spaces and how individual user decisions are constrained. For example, someone opting out of Second Life, and creating their own "Third Life," might well find that none of his friends follow him to his own virtual world, and that the creators of Second Life sue for copyright and trademark infringement to the extent the newer virtual world mimics their own. Those trying to defect to the alternative social networks Google+ and Ello have experienced this coordination problem directly: maybe they and some enterprising friends establish a presence there, only to find that ninety percent of the rest of their social network is too busy or uninterested to join them. Technical and practical challenges to creating a user experience sufficiently similar to attract users of the Second Life interface, while sufficiently different to avoid infringing intellectual property, may well be insuperable.

To deal with these dynamics, Professor Zarsky has recognized that

---

6. *See generally* Danielle Keats Citron, *Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age*, 80 S. CAL. L. REV. 241 (2007).

7. Tal Z. Zarsky, *Privacy and Data Collection in Virtual Worlds*, *in* THE STATE OF PLAY: LAW, GAMES, AND VIRTUAL WORLDS 217 (Jack M. Balkin & Beth Simone Noveck eds., 2006).

8. Bernard E. Harcourt, *Digital Security in the Expository Society: Spectacle, Surveillance, and Exhibition in the Neoliberal Age of Big Data* 13 (Columbia Law Sch. Pub. Law & Legal Theory Working Paper Grp., Paper No. 14-404, 2014), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2455223; *see also* David Gray & Danielle Keats Citron, *The Right to Quantitative Privacy*, 98 MINN. L. REV. 62 (2013).

TOSs and EULAs are, by and large, contracts of adhesion, and that only a public law solution can address the imbalances inevitable in these contracts, which in most cases amount to little more than private legislation of terms by the dominant party. We agree. Regulatory initiatives are essential to guard consumers who can hardly anticipate all potential uses of data on their own.

Professor Zarsky's attention to threats to the sensitive online ecology of social software continued in his work *Law and Online Social Networks*.[9] He analyzed the positive and negative social effects of interactions in these environments. Professor Zarsky is one of the first scholars to notice the importance of gaming in social networks—efforts to manipulate or fake the bonds of trust that create social capital and cooperation. We believe that, unregulated, scoring could lead to exactly the same issues: people pervasively manipulating their own identities to gain monetary or other advantages.[10]

Professor Zarsky has balanced two sets of competing demands on legal scholars in technologically cutting-edge fields. While a principle of subsidiarity recommends market- and contract-based remedies for many wrongs that can happen on these networks, problems of "astroturf" (*i.e.*, commodified and artificial support) and manipulation can probably only be addressed via dedicated entities with the technical expertise to patrol against them. One must simultaneously understand and engage with new technical developments unanticipated by lawmakers, and try to identify the issues that will recur as future developments supersede present controversies. Professor Zarsky masterfully balanced these imperatives in *Law and Online Social Networks*, both comprehending the new opportunities for distributed information creation generated by social networks, and isolating the core issues of gaming and authentication that will prove nettlesome in virtually any foreseeable instantiation of social software. The law of consumer scoring still has to grapple with both issues.[11]

As reflected in his response to us, Professor Zarsky recognizes the limits of legal solutions to the challenges of the scored society. In his

---

9. Tal Z. Zarsky, *Law and Online Social Networks: Mapping the Challenges and Promises of User-Generated Information Flows*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 741 (2008).

10. Frank Pasquale, *Facebook's Model Users*, NEW CRITICALS (July 6, 2014), http://www.newcriticals.com/facebooks-model-users/print.

11. PAM DIXON & ROBERT GELLMAN, WORLD PRIVACY FORUM, THE SCORING OF AMERICA: HOW SECRET CONSUMER SCORES THREATEN YOUR PRIVACY AND YOUR FUTURE (2014), *available at* http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf.

reflections on anonymity, transparency, and pseudonymity in *Thinking Outside the Box*,[12] he meticulously detailed the inevitable tradeoffs that occur online in tracking individual action—tradeoffs policymakers will have to consider as they weigh the virtues and vices of requiring better data collection practices by scorers. With this appreciation of Professor Zarsky's work, we have now set the stage for responding on mutually agreeable normative foundations to the incisive analysis he has offered to enrich our own.

I. EQUAL OPPORTUNITY IS IMPERILED DUE TO THE SCORING OF POLITICAL EXPRESSION

In his response, Professor Zarsky focuses on how scoring may have a negative impact on traditionally protected groups, such as racial minorities.[13] He observes that, at least in the United States, the Supreme Court is not disposed to include many more such groups under the protection of the Fourteenth Amendment. Existing state constitutional law and statutory law, however, is not so confined. Civil rights protections can and do reach newly recognized disadvantaged groups, including sexual minorities.[14] We look forward to learning more about how other countries might take a more flexible approach.

However, our concerns about scoring extend beyond the protections afforded traditionally disadvantaged groups from a constitutional or statutory perspective. The key to understanding the menace of scoring in a modern, Big Data economy is the volume, velocity, and variety of information that could be fed into a score.[15] The legal academy is still catching up to the shocking empirical findings of security researchers, privacy law scholars, and computer scientists. To take only one example, Pam Dixon and Robert Gellman have unearthed thousands of scoring

---

12. Tal Z. Zarsky, *Thinking Outside the Box: Considering Transparency, Anonymity, and Pseudonymity as Overall Solutions to the Problems in Information Privacy in the Internet Society*, 58 U. MIAMI L. REV. 991 (2004).

13. For other thoughtful work on the potential for antidiscrimination law in troubling uses of Big Data, see Solon Barocas & Andrew D. Selbst, Big Data's Disparate Impact (Oct. 19, 2014) (unpublished), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899.

14. Some states extend civil rights protections not only to racial and religious minorities but also to people targeted due to their gender and sexual orientation. *See, e.g.*, Kathleen W. Mikkelson, *California's Civil and Criminal Laws Pertaining to Hate Crimes*, OFF. OF ATT'Y GEN. OF CAL., http://oag.ca.gov/civil/htm/laws (updated Feb. 25, 1999).

15. *See* EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 4 (2014), *available at* http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf.

systems. Outsiders have no idea what is going into many of them.[16] And yet these scores could become decisive in contexts ranging from employment to insurance to marketing and beyond.

Our article was an effort to challenge the presumption of benign, targeted scoring. The Big Data economy is premised on the accumulation of massive amounts of data, and it is all but certain that some of it will be sensitive data or will be pieced together to make derogatory inferences. Just as the ever-declining price of surveillance and the pervasiveness of sensor networks have revolutionized Fourth Amendment law, the very cheapness of data requires us to rethink privacy law. Professor Zarsky mentions the possibility that there are already lists of Muslim and Jewish individuals in commercial databases. Add to that lists of gay individuals, Democrats, Republicans, Socialist Party supporters, breast cancer survivors, Alcohol Anonymous participants, fans of mixed martial arts, Krav Maga members, violent video game addicts, and literally millions of other categories.

It is one thing to go through life with a sense that authorities may be able to scrutinize all of one's data in the context of criminal investigations or intelligence gathering. We addressed the deep concerns such surveillance raises in our article *Network Accountability for the Domestic Intelligence Apparatus*[17]—a piece that the Snowden revelations now show to be prescient in terms of the dark possibilities it raised but that we could not entirely confirm in 2011. With the leaks of governmental insiders, we have proof that pervasive, continuous, and totalizing surveillance is the order of the day.

That said, it is quite another matter, and in many respects the more chilling extension of surveillance, to understand that one's data is being processed in literally thousands of scores that cannot be reviewed, understood, or challenged. "Do I dare retweet the #Ferguson hashtag, lest some future employer score that as an indicator of rebelliousness?" a person may reasonably ask herself. "Are there risks in calling Edward Snowden a whistleblower, lest that suggest an anti-government agenda?" is another question that is reasonable for social media participants to ask themselves.

Until we have much better knowledge of scoring practices, and quite possibly until we have law explicitly restricting employers from basing hiring, firing, and promotion decisions on generalized assumptions based

16. DIXON & GELLMAN, *supra* note 11, at 7.

17. Danielle Keats Citron & Frank Pasquale, *Network Accountability for the Domestic Intelligence Apparatus*, 62 HASTINGS L.J. 1441 (2010–2011).

on political expression, a troubling burden on expressive freedom will persist.

Whatever the current state of First Amendment doctrine, a regime of total surveillance undermines the free development of personality upon which free expression depends.[18] The power to watch is the power to attack, embarrass, and destroy reputations. As Professor Daniel J. Solove shows, privacy is not just a problem of concealing isolated facts.[19] Of great concern is the collection and analysis of a critical mass of data. Our lives are starting to become an open book for those powerful or rich enough to score our profiles.

We need to think of privacy as being as much a vindication of our First Amendment as our Fourth and Fifth Amendment rights.[20] Professor Marc Jonathan Blitz has explored the intersection of free speech and privacy values.[21] Will individuals hesitate to join mental illness support groups on Facebook, once they are aware that an ever-growing array of body or mind scores may be used against her? Will they refrain from "liking" fringe political groups on Facebook, once they realize that their affiliations on social media are ending up in scores that can have a detriment on their careers?

The technological tools for matching digital records are staggering. State restrictions on the use of that data (and scores based on it) can be an important step toward giving individuals a chance to form and express opinions and affiliations in peace—without fearing an endlessly ramifying series of classifications made and opportunities possibly denied, on account of faceless and secretive data miners.

II. PRIVACY AND POSITIVE-SUM INNOVATION: MUTUALLY REINFORCING GOALS

A balanced and thoughtful reconciliation of the interests of data brokers, data subjects, scorers, and users of data and scores is important. We do not want to unduly burden a nascent industry. But we should also realize that privacy and innovation are mutually reinforcing constructs when, as in our case, critical aspects of privacy protection require the

---

18. *See* JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE (2012) (on dangers of modulation).

19. DANIEL J. SOLOVE, NOTHING TO HIDE: THE FALSE TRADEOFF BETWEEN PRIVACY AND SECURITY (2011).

20. *See* Daniel J. Solove, *The First Amendment as Criminal Procedure*, 112 N.Y.U. L. REV. 112, 114–15 (2007).

21. Marc Jonathan Blitz, Stanley *in Cyberspace: Why the Privacy Protection of the First Amendment Should Be More Like That of the Fourth*, 62 HASTINGS L.J. 357, 359 (2010–2011).

validation of the data used. In reliable Big Data science, researchers invest a great deal of time and effort in cleaning up data, assuring that it is actually accurate and verifiable.[22] In commercial contexts where opportunities and livelihoods are at stake, the case for assuring data integrity applies *a fortiori.*

Moreover, there is now an opportunity to shape scoring systems toward positive-sum innovation, as opposed to ever more baroque strategies of discrimination. Pam Dixon has accused firms of using "vulnerability-based marketing" to target consumers.[23] In one disturbing example, marketers were urged to place ads at times of the day when women felt worst about themselves.[24] In another, consumers were unaware that lead generators for credit were selling their names to the "highest bidder," including firms more than ready to charge near-usurious interest rates.[25] The U.S. Public Interest Research Group (U.S. PIRG) correctly argues:

> [S]ome of the non-transparent, deceptive pages you are led to on the Internet when you type "I need a loan," may appear to be lenders, but aren't. These websites are actually "lead generators," that ask you a few questions to determine your value and then auction you off to the highest bidder, often an online payday lender or for-profit school. Lead generators are the target of numerous enforcement inquiries, including by New York. . . . [I]f the protections offered by a regulated prescreening system for financial marketing are diluted by a switch to scores generated using largely unregulated Internet algorithms created through the sharing of cookies and all these other tracking bits between and among a vast interconnected network of business-to-business firms that consumers don't know about or do business with, consumers will be harmed.[26]

---

22. Steve Lohr, *For Data Scientists, "Janitor Work" Is Hurdle to Insights*, N.Y. TIMES, Aug. 18, 2014, at B4.

23. *See WPF's Data Broker Testimony Results in New Congressional Letters to Data Brokers about Vulnerability-Based Marketing*, WORLD PRIVACY F. (Feb. 3, 2014), http://www.worldprivacyforum.org/2014/02/wpfs-data-broker-testimony-results-in-new-congressional-letters-to-data-brokers-regarding-vulnerability-based-marketing/.

24. Lucia Moses, *Marketers Should Take Note of When Women Feel Least Attractive: What Messages to Convey and When to Send Them*, ADWEEK (Oct. 2, 2013, 6:44 AM), http://www.adweek.com/news/advertising-branding/marketers-should-take-note-when-women-feel-least-attractive-152753.

25. Ed Mierzwinski, *We Join FTC Event on Big Data E-Scores*, U.S. PIRG EDUC. FUND (Mar. 19, 2014), http://www.uspirgedfund.org/blogs/eds-blog/usp/we-join-ftc-event-big-data-e-scores.

26. *Id*.; *see also* Ed Mierzwinski & Jeff Chester, *Selling Consumers Not Lists: The New World of Digital Decision-Making and the Role of the Fair Credit Reporting Act*, 46 SUFFOLK L. REV. 845,

The National Consumer Law Center has concluded that, at least so far, Big Data has been a "big disappointment" for scoring creditworthiness.[27]

We also should be worried about misdirection of the innovation of scoring in the employment context—particularly if firms can effectively hide misconduct via scores. Existing laws prohibit some discriminatory uses of the data. For example, an employer cannot fire workers simply because they have an illness. But Big Data methods are able to predict diabetes from a totally innocuous data set (including items like eating habits, drugstore visits, magazine subscriptions, and the like). And the analyst involved, whether inside or outside the firm, could easily mask the use of health-predictive information.

For example, a firm could conclude a worker is likely to be diabetic and that she is likely to be a "high cost worker" given the significant monthly costs of diabetic medical care. Given the proprietary nature of the information involved, the most the firm is going to tell the fired (or unhired) worker is the end result: the data predicted that her cost to the firm was likely to be greater than the value she produced. Most of the time, companies need not even offer that rationale. Unexplained and unchallengeable, Big Data becomes a star chamber.

We do not have to put up with this stigmatic profiling. State legislatures should require employers to reveal to employees *all the databases* of information used to make judgments about them. If we do not get that access, we may never know why key decisions are made. And secrecy is a discriminator's best friend: unknown unfairness can never be challenged, let alone corrected.[28] Without mandating privacy-respecting innovation, new technology can be abused in order to hide (and ultimately promote) discrimination, rather than to promote truly productive innovation.

---

846 (2013).

27. NAT'L CONSUMER LAW CTR., BIG DATA: A BIG DISAPPOINTMENT FOR SCORING CONSUMER CREDIT RISK (2014), *available at* http://www.nclc.org/issues/big-data.html.

28. It is already difficult to challenge discrimination in hiring practices, for example. The Supreme Court recently granted certiorari in a case where Abercrombie & Fitch refused to hire a young Muslim woman because her hijab did not comport with the company's "Look Policy." The Tenth Circuit held that even though the woman would have been entitled to a religious accommodation under Title VII, the company's decision not to hire her did not violate the law because she never provided explicit notice that she would require such an accommodation. *See* Equal Emp't Opportunity Comm'n v. Abercrombie & Fitch Stores, Inc., 731 F.3d 1106 (10th Cir. 2014), *cert. granted*, 83 U.S.L.W. 3089 (U.S. Oct. 2, 2014) (No. 14-86).

III. CALIBRATING A POLICY RESPONSE

Professor Zarsky is correct to say that the scope and level of protections afforded to individuals should vary depending on the degree of harm suffered. He raises important empirical questions about the level of harm that we should expect in areas like marketing and insurance. The policy response should also be calibrated with respect to characteristics of the scoring systems at issue.

Consider, for instance, the following chart:

	Opaque Data, Algorithms, & Outputs	**Opaque Data & Algorithms, Transparent Outputs**	**Opaque Data, Transparent Algorithms & Outputs**	**Transparent Data, Algorithms, & Outputs**
Unrevisable Judgment, Unreformable System	**1**	**2**	**3**	**4**
Revisable Judgment, Unreformable System	**5**	**6**	**7**	**8**
Unrevisable Judgment, Reformable System	**9**	**10**	**11**	**12**
Revisable Judgment, Reformable System	**13**	**14**	**15**	**16**

The primary targets for legal reform should be the systems associated with the boxes in the upper, left-hand corner in the chart above. The lower, right-hand corner brings us closer to a zone of technological due process. There is a far greater chance of competition in scenarios 4, 8, 12, and 16 than when some critical aspect of the scoring process is kept secret. But even if multiple open systems are competing, we should still have some concerns if, as in scenario 4, they generate unrevisable judgments, and refuse to open themselves up to the possibility of reform

in response to the concerns of scored individuals.

IV. PROBLEMATIC ASSUMPTIONS

We should also note that Professor Zarsky does assume away some of the problems that worried us.[29] He assumes that scores "are premised upon the individuals' previous behaviors, rely upon non-spurious correlations between individual attributes and problematic behaviors the process is trying to predict. (A default, a risk, or poor work performance are some key examples)."[30]

However, we are concerned about many situations where scores are premised on false or distorted accounts of individual behavior. Consider two different examples. Let's first assess credit reports. Millions of Americans have errors on their credit reports. Credit reporting is one of the most highly regulated, perhaps the most highly regulated, data gathering used for scoring. In nearly all other types of scoring, individuals do not know that the scoring is done, how the data is gathered, what data is fed into the relevant algorithms, the nature of these algorithms, or the effect of the scoring.[31]

Now, let's turn to analytics firms that crunch data to search for and assess talent in particular fields.[32] Remarkable Hire scores a job candidate's talents by looking at how others rate his or her online contributions.[33] Talent Bin and Gild create lists of potential hires based on online data.[34] Big-name companies like Facebook, Wal-Mart, and Amazon use these technologies to find and recruit job candidates.[35] Will algorithms give high scores to individuals who have been harassed online with defamation, threats, and the posting of nude photos that have either been stolen from their online accounts or exposed in violation of their trust? Will they identify harassed individuals as top picks for employment if those targeted individuals have withdrawn from online life? Will they discount online abuse so that victims can be evaluated on their merits rather than the falsehoods and privacy invasions spread by their harassers? One can only guess the answers to these questions, but

---

29. Zarsky, *supra* note 1, at 1383 ("[T]he following discussion is premised upon several non-trivial assumptions regarding the scoring process.").

30. *Id.*

31. *See* DIXON & GELLMAN, *supra* note 11.

32. Matt Richtel, *I Was Discovered By An Algorithm*, N.Y. TIMES, April 28, 2013, at BU1.

33. *Id.*

34. *Id.*

35. *Id.*

our bet is that the falsehoods will hold sway.

Thus, one of our key policy proposals: those harboring significant amounts of data ought to have some certified indication of its provenance. Such certifications should be regularly audited. Data should not be allowed to persist without certification of its provenance and accuracy. Until those types of protection are in place, it is in the state's interest to tightly regulate the transfer of health data, much of which the state itself required to be created.

Professor Zarsky also assumes that "the scoring schemes structured by statisticians in the back office are indeed followed to a tee by those in the field."[36] Unfortunately, if the U.S. experience is any indication, precisely the opposite may often be the case. As one finance expert has observed, "the more complex the algorithm, the more opportunities it provides to the salespeople to 'game' and arbitrage the system in order to commit fraud."[37] Promoted as a road to opportunity, the aspiration to price credit according to scores has had a darker side. Abuses quickly piled up, as "some large financial institutions peddled mortgages to people who could not possibly pay the monthly rates."[38] Subprime-structured finance generated enormous fees for middlemen and those with "big short" positions, while delivering financial ruin to many end-purchasers of mortgage-backed securities and millions of homebuyers.[39]

In conclusion, we are happy to have had this opportunity to further develop and clarify our views in response to Professor Zarsky. His work has inspired important research in cyberlaw. We take his recommendations seriously as we and other scholars pursue a research agenda to address Big Data's perils for disadvantaged groups.

---

36. Zarsky, *supra* note 1, at 1384.

37. Ashwin Parameswaran, *How to Commit Fraud and Get Away With It: A Guide for CEOs*, MACRORESILIENCE (Dec. 4, 2013, 4:19 PM), http://www.macroresilience.com/2013/12/04/how-to-commit-fraud-and-get-away-with-it-a-guide-for-ceos/.

38. MARGARET ATWOOD, PAYBACK: DEBT AND THE SHADOW SIDE OF WEALTH 8 (2008).

39. *See* MICHAEL LEWIS, THE BIG SHORT (2010); JENNIFER TAUB, OTHER PEOPLE'S HOUSES: HOW DECADES OF BAILOUTS, CAPTIVE REGULATORS, AND TOXIC BANKERS MADE HOME MORTGAGES A THRILLING BUSINESS (2014); Robert Brenner, *What Is Good for Goldman Sachs Is Good for America: The Origins of the Current Crisis* (2009), *available at* http://www.escholarship.org/uc/item/0sg0782h.